Assumptions

Robert Howard

10 January 1996

Some modeling terms ...

Assumptions -
Conditions or suppositions that are posited as true.

Constraints -
Generally anything that constrains the implementation of the model (e.g., the actual design of a specific subsystem or device).

Basis Functions -
Function that describes the operation or response of an object within the model (e.g., statistics of stochastic loads, equations that calculate hardware response times).

Modeling Parameters -
Parameters (constants) that are entered into the model (e.g., hardware characteristics).

Road map

Assumption Summaries

- **System assumptions**
- **Input assumptions**
- **Distribution assumptions**
- **Archive assumptions**
- **Processing assumptions**
- **Reprocessing assumptions**
- **Modeling assumptions**

Note: the summaries are indexed to entries in the Assumptions Appendix. Each entry in the appendix gives additional information on the assumption.

System Assumption Summary

1. **Product and processing load and growth are as specified by the AHWGP.**

2. **Processing load and file sizes required for the creation of browse products within ECS are included in the AHWGP data.**

3. **EDOS has a single interface with each DAAC that receives Level 0 data. The data on that interface consists of only the data which is destined for that DAAC.**

4. **At DAACs that do not operate 24 hours/day, 7 days/week, the DMS and DSS operate autonomously 24 hours/day throughout the week. This means that user query and electronic data distribution are allowed 24 hours/day, 7 days/week.**

Input Assumption Summary

1. Non-ECS products needed for ECS product dependencies arrive at ECS on a periodic basis (e.g., daily, hourly, etc.).

2. A specific step to create Level 0 granules is not required. Instead, the PGE that uses the Level 0 data performs the granulization internal to itself.

3. Data for V0 migration are blocked into large groups (sets of files are passed together into ECS) and delivered electronically.

4. Data regularly ingested by ECS from other non-ECS data generators (e.g., Landsat, ASF, TSDIS) are blocked into large groups of files and delivered electronically.

Input Assumption Summary (cont).

5. ASTER Level 1A and 1B data arrives on physical media once per day.

6. ASTER data received on physical media lags its spacecraft acquisition time by 7 days.

7. The tape containing ASTER L1A and L1B is processed during third shift.

8. Data received from EDOS lags its spacecraft acquisition time by 21 hours.

9. V0 client access (V0 into ECS) is negligible in 3rd Qtr 1999.

Distribution Assumption Summary

1. Coordinated delivery of products from multiple DAACs within a single order is not required.

2. All sites have the same standard distribution devices (media).

3. The volume of data distributed from a DAAC to the users is 2 times the amount inserted into the DAAC's archive. This is also known as the 2x volume distribution assumption.

4. Data for user electronic "pull" is held on distribution disk 24 hours before it is deleted.

Distribution Assumption Summary (cont.)

5. All data requested for user electronic “pull” is picked up.

Archive Assumption Summary

1. Both subsetting and user-specified processing operate on data that is available from the archive.
2. The data output to the archive is time-order blocked to tape size.
3. Other than that imposed by the blocking, there is no organization of data on the archive data tapes.
4. All sites use the same type of Archive Hardware.
5. The 30-day rolling store products at ASF are held in the archive on reusable media.

Processing Assumption Summary

1. Effective CPU processing power is 25% of manufacturer's peak-rated CPU power.

2. Individual processors within DPS are not timeshared.

Reprocessing Assumption Summary

1. **The reprocessing load is twice the average load caused by “first-time” processing, subject to phasing capacity.**
    - **Phasing capacity in 3rd Qtr 1999 (epoch k) is 2.2**
        - **This allows for first-time processing and x1 reprocessing**

2. **Reprocessing is done using the “head-of-chain” paradigm.**

Modeling Assumption Summary

1. The model does not keep track of any correlation between user requests.

2. Network throughput is determined by network characteristics and traffic load, not by the characteristics of the devices in use at the source and sink nodes.

Model Basis Functions

10 January 1996

Basis Functions Summary

1. Number of User Requests
2. DAAC to which request is directed.
3. Volume of data per request for ad hoc requests.
4. Service Request type.
5. Processing load due to manipulate requests.
6. Number of files input (from archive) during a manipulate.
7. Decimation ratio
8. Portion of data that is distributed electronically.

Note: the summaries are indexed to entries in the Basis Functions Appendix. Each entry in the appendix gives additional information on the assumption.

Basis Functions Summary (cont.)

9. Portion of electronic distribution that is “pulled” (as opposed to “pushed”).
10. CPU processing time for each PGE.
11. I/O time for each PGE.
12. Elapsed processor time (also sometimes called “wall clock time”).
13. Archive read/write time.
14. Exclusive (switched, non-shared) Network transfer time.
15. Shared Network transfer time.
16. Algorithm to retrieve from or insert a file into the archive hardware.
17. Archive Tape Fetch/Replace Time.
18. Archive Tape Search Time.

Basis Functions Summary (cont.)

19. Archive Tape Rewind Time.
20. Probability that a standard product granule (file) is subject to a subscription (standing order).
21. Job scheduling algorithm.
22. Time data is held on disk at DPS.
23. Time data is held on disk at DSS.
24. Type of client access.
25. Percentage of requests referred to the V0 gateway.
26. Single/Multiple DAAC query ratio.

Model Parameters

10 January 1996

Model Parameters (1)

Archive Operations (currently Bosch silo with D3 media is being used)

Operation	STK,D3	Bosch,D3	STK,NTP	Bosch,NTP	Units
Fetch	7.80	25.42	7.80	25.42	sec.
Mount	17.00	17.00	20.00	20.00	sec.
Search	106.00	106.00	60.00	60.00	sec.
Read/Write	8.40	8.40	9.00	9.00	MB/sec
Rewind	85.00	85.00	60.00	60.00	sec.
Dismount	8.50	8.50	10.00	10.00	sec.
Replace	7.80	25.42	7.80	25.42	sec.
Cmd/Ack Delay	16.00	16.00	16.00	16.00	sec.
FSMS Delay	12.00	12.00	12.00	12.00	sec.
Capacity	50.00	50.00	10.00	10.00	GB

Source: DSS Staff; from testbed measurements and derated vendor-supplied information.

Model Parameters (2)

DAAC Hours of Operation (3rd Qtr 1999 / epoch k):

- **ASF 8x5**
- **EDC 24x7**
- **GSFC 24x7**
- **JPL 8x5**
- **LaRC 24x7**
- **NSIDC 8x7**
- **ORNL 8x5**

Note: Also see System Assumption (4).

Model Parameters (3)

Priorities (weighting):

• Large Request	25
• Reprocessing	33
• Users	50
• First-time processing	100

Note: Higher numbers imply higher priority

Model Parameters (4)

Network Throughput (will be derated by requesting organization)

- **HiPPI 100 MB/second**
- **FDDI 12.5 MB/second**
- **Ethernet 1.5 MB/second**

Latency$_{disk}$	**0.024 seconds per block**
Block Size$_{disk}$	**500 kBytes**
Bandwidth$_{I/O}$	**50 MBytes/second**

Model Parameters (5)

Estimated Peak Processor Power (before derating)

- **SGI (Epoch e)   360 MFLOPS**
- **SGI (Epoch k)   720 MFLOPS**
- **DEC Alpha   200 MFLOPS**
- **Cray   900 MFLOPS**

Notes:

- **Currently the model is run with only SGI machines**
  - **except for DAO**
- **DAO is run with a small number of extremely large machines**
  - **DAO will acquire the most processing power within budget limitations**
  - **Model handles DAO as length of time their processors are busy**
- **Data supplied by vendors.**

Assumptions Appendix

System Assumption (1)

Assumption: Product and processing load and growth are as specified by the AHWGP.

Rationale: The AHWGP is the best source of data on algorithm performance. Also derived from Technical Direction #11.

Impact: Resource requirements would be larger or smaller depending on whether the predicted loads are higher or lower.

Notes: None.

System Assumption (2)

Assumption: Processing load and file sizes required for the creation of browse products within ECS are included in the AHWGP data.

Rationale: A few browse products are listed as separate PGEs in AHWGP data.

Impact: Additional processing power needed at the DAACs to create the browse products. Additional archive space and DMS entries needed for browse products. Also additional working storage needed at the data server.

Notes: This assumption does not address browse products created at an SCF.

System Assumption (3)

Assumption: EDOS has a single interface with each DAAC that receives Level 0 data. The data on that interface consists of only the data which is destined for that DAAC.

Rationale: Current understanding of the EDOS Ops Concept.

Impact: If this is not the case, buffer (disk) sizes may be underestimated. There may also be changes in communication costs.

Notes: ASTER, Landsat, and ASF data are not supplied to ECS via EDOS.

System Assumption (4)

Assumption: At DAACs that do not operate 24 hours/day, 7 days/week, the DMS and DSS operate autonomously 24 hours/day throughout the week. This means that user query and electronic data distribution are allowed 24 hours/day, 7 days/week.

Rationale: Current M&O ops concept.

Impact: If the DMS and DSS are not operating 24 hours per day, 365 days per year, those elements will have to be resized (upwards) in order to work off the daily and/or weekly backlogs.

Notes: This assumption implies that generation of standard data products and physical distribution media creation do not occur during the off hours.

Input Assumption (1)

Assumption: Non-ECS products needed for ECS product dependencies arrive at ECS on a periodic basis (e.g., daily, hourly, etc.).

Rationale: For a standard product to be produced regularly, ECS must be assured that its inputs are available in a timely fashion.

Impact: Additional staging of data files may be required, with subsequent increases in archive read/write stations and robotics, LANs, and on-line storage.

Notes: For some data sources, such as the NOAA weather models, this assumption is manifestly valid.

Input Assumption (2)

Assumption: A specific step to create Level 0 granules is not required. Instead, the PGE that uses the Level 0 data performs the granulization internal to itself.

Rationale: Current design of ECS and its toolkit.

Impact: If a granulization step has to be added, PGEs to accomplish this must be added to the baseline. This will increase the scheduler and processing loads.

Notes: ASTER Level 1A and 1B are generated outside of ECS, therefore it is assumed that the granulization is also done outside of ECS.

Input Assumption (3)

Assumption: Data for V0 migration are blocked into large groups (sets of files are passed together into ECS) and delivered electronically.

Rationale: Current design ops concept.

Impact: Little, since Data Server blocking buffer will perform the same function.

Notes:

- Migration of some V0 data via physical media is being evaluated by the design teams.
- This assumption avoids the necessity of building a data generator in the dynamic model for each different data set in the V0 system.

Input Assumption (4)

Assumption: Data regularly ingested by ECS from other non-ECS data generators (e.g., Landsat, ASF, TSDIS) are blocked into large groups of files and delivered electronically.

Rationale: Current design ops concept.

Impact: Little, since Data Server blocking buffer will perform the same function.

Notes: This assumption avoids the necessity of building a data generator in the dynamic model for each different data set.

Input Assumption (5)

Assumption: ASTER Level 1A and 1B data arrives on physical media once per day.

Rationale: Interface with the Japanese data system (ASTER IRD and PIP).

Impact: If this interface is changed, ECS will have to resize the appropriate interface to account for the added data traffic.

Notes: None

Input Assumption (6)

Assumption: ASTER data received on physical media lags its spacecraft acquisition time by 7 days.

Rationale: Interface with the Japanese data system (ASTER IRD and PIP).

Impact: Little to none, unless the data is received on the day the data was acquired. The reason for this is that the data needed for input to the ASTER processes will have to be retrieved from the archive.

Notes: None.

Input Assumption (7)

Assumption: The tape containing ASTER L1A and L1B is processed during third shift.

Rationale: Processing in non-prime time minimizes the impact on users.

Impact: Little anticipated, some change in duties of M&O personnel between first and third shifts.

Notes: None.

Input Assumption (8)

Assumption: Data received from EDOS lags its spacecraft acquisition time by 21 hours.

Rationale: By requirement, EDOS has up to 21 hours to process its inputs.

Impact: Small.

Notes: None.

Input Assumption (9)

Assumption: V0 client access (V0 into ECS) is negligible in 3rd Qtr 1999.

Rationale: By that time most users who wish to use ECS will have transitioned into using ECS directly (as opposed to using V0 to access ECS services).

Impact: Small.

Notes: This does not imply that ECS users will not access V0 data. For ECS access into V0, see Basis Functions (25).

Distribution Assumption (1)

Assumption: Coordinated delivery of products from multiple DAACs within a single order is not required.

Rationale: Lowers ECS costs and follows current ECS design.

Impact: Increased costs due to additional storage needed for DAAC-to-DAAC transfers. Increased network costs due to additional DAAC-to-DAAC traffic. Increased coordination between DAACs to fill orders.

Notes: This does not require users receiving data via electronic “pull” to know where (which DAACs) they need to go to pull data from disk. The client software may determine which DAAC via the URL(s).

Distribution Assumption (2)

Assumption: All sites have the same standard distribution devices (media).

Rationale: Technical Requirement (DADS2490).

Impact: Some (relatively small) reprogramming of the model.

Notes: While every site has the same type of hardware, different sites may have differing number of read/write stations.

Distribution Assumption (3)

Assumption: The volume of data distributed from a DAAC to the users is 2 times the amount inserted into the DAAC's archive. This is also known as the 2x volume distribution assumption.

Rationale: Technical direction by ESDIS.

Impact: A larger volume of data distribution would result in an increased amount of archive hardware (robotics and read/write stations) as well as distribution hardware (working storage and media hardware).

Notes: The interpretation of the technical direction is that the 2x is applied on a per DAAC basis.

Distribution Assumption (4)

Assumption: Data for user electronic “pull” is held on distribution disk 24 hours before it is deleted.

Rationale: While it is likely that users will pick up their data before the 24 hours have elapsed, to model this would require a distribution of pick up times. This distribution is unavailable, so the worst case assumption (hold for 24 hours) is made.

Impact: If the data is held longer on disk, either more storage will be required or the time the data is held on the disk will have to be decreased. Conversely, if the users pick their data before 24 hours, either less disk will be required or the maximum “hold time” can be increased beyond 24 hours.

Notes: In practice, the length of time the data will be retained on disk will be set by each DAAC.

Distribution Assumption (5)

Assumption: All data requested for user electronic "pull" is picked up.

Rationale: A user can register a subscription specifying user electronic "pull". It is possible that the user is busy or out-of-town and does not pick that data up before it is removed from disk. He would then request the data a second time. It is assumed that this happens infrequently.

Impact: If data is requested a second time by the same user, the number of robotics and read/write stations at the archive would have to be increased.

Notes: See Distribution Assumptions (4).

Archive Assumption (1)

Assumption: Both subsetting and user-specified processing operate on data that is available from the archive.

Rationale: ECS is designed to facilitate EOS research. If a user is submitting a non-EOS algorithm which neither reads nor writes EOS data, then ECS is being used only as a back-end processing server. There are other facilities that are designed to offer this service.

Impact: Allowing such requests would increase processing and storage requirements. The amount of increase could be anywhere from slight to large due to varying nature of this type user requests .

Notes: None.

Archive Assumption (2)

Assumption: The data output to the archive is time-order blocked to the tape size.

Rationale: Current data server ops concept. Essentially what happens is data to be archived is held in a buffer (either disk or other media) until there is enough data to fill an archive tape. When enough data has arrived, all the files are transferred to tape. After the files have been written to tape they are removed from the buffer.

Impact: Not using blocking or a related technique would result in more accesses to the archive. This would result in a substantial increase in the number of robotics and read/write stations required. Blocking does require additional storage to implement the buffer.

Notes: None.

Archive Assumption (3)

Assumption: Other than that imposed by the blocking, there is no organization of data on the archive data tapes.

Rationale: Current data server ops concept.

Impact: Depending on user and production access patterns, tape organization may lower the number of robotics and read/write stations required. Additionally, access times may also be lowered. Organizing the data is not without its own resource and access time costs, however.

Notes: Examples of organizing data on tape would be a single data set on per tape or a limited geophysical location (but multiple data sets) per tape.

Archive Assumption (4)

Assumption: All sites use the same type of Archive Hardware.

Rationale: Simplifies the dynamic model.

Impact: Some (relatively small) reprogramming of the model.

Notes: While every site has the same type of hardware, different sites may have differing number of robots and/or read/write stations.

Archive Assumption (5)

Assumption: The 30-day rolling store products at ASF are held in the archive on reusable media.

Rationale: Current DAAC-specific design (see DID 305).

Impact: If another device is selected as the mechanism for 30-day products it will have to be sized and costed.

Notes: None.

Processing Assumption (1)

Assumption: Effective CPU processing power is 25% of manufacturer's peak-rated CPU power.

Rationale: Technical Requirement (PGS-1301).

Impact: If the effective power tends toward the peak-rated power, then fewer processors would be needed.

Notes: The 25% derating is consistent with Dr. Dongarra's (ORNL) benchmarking of multiple computer systems.

Processing Assumption (2)

Assumption: Individual processors within DPS are not timeshared.

Rationale: For compute-bound processes, timesharing decreases the processor efficiency.

Impact: If a significant number of PGEs are I/O-bound, then the current approach may undersize the number of CPUs and disks that are needed.

Notes: Once a CPU is allocated to a PGE, it is dedicated to that PGE until it terminates. Multiple CPUs may be running simultaneously on different PGEs or different instantiations of the same PGE.

Reprocessing Assumption (1)

Assumption: The reprocessing load is twice the average load caused by "first-time" processing, subject to phasing capacity.

Rationale: Technical Requirement (PGS-1300).

Impact: Requirement PGS-1300 essentially sets aside a capacity for reprocessing. How this capacity is applied is decision made by the individual DAACs. Therefore, the impact of this assumption can not be determined until DAAC policy is set.

Notes: Reprocessing Assumption (2) interprets how "twice" in this assumption is to be applied.

Reprocessing Assumption (2)

Assumption: Reprocessing is done using the "head-of-chain" paradigm.

Rationale: Use of the head-of-chain paradigm lowers the resources required to perform reprocessing.

Impact: Use of another paradigm will cause increased (sometimes severely increased) requirements for disk storage, networks, and archive hardware.

Notes: The various paradigms and a brief analysis of their impact was presented at IDR-B.

Modeling Assumption (1)

Assumption: The model does not keep track of any correlation between user requests.

Rationale: It is difficult to determine whether user requests are correlated or not, and if correlated, what the correlation is. In the absence of this determination, the more conservative assumption (this one) should be used.

Impact: If user requests are correlated, there may be some savings in terms of amount of robotics needed.

Notes: The practical impact of this assumption is that the dynamic model always assumes that each request will require mounting a different tape than the one that is already on the read/write station.

Modeling Assumption (2)

Assumption: Network throughput is determined by network characteristics and traffic load, not by the characteristics of the devices in use at the source and sink nodes.

Rationale: This must be the case if networks are to be efficiently used. If peripherals become the pacing item, then high-speed networks will be idle waiting for the peripheral devices.

Impact: If peripherals become the pacing item, then it would be more cost effective to buy many low-speed networks.

Notes: While throughput and utilization are determined by network characteristics and traffic load, elapsed time before a file is available is dependent on the slowest link.

Basis Functions Appendix

Basis Functions (1)

Parameter: Number of User Requests

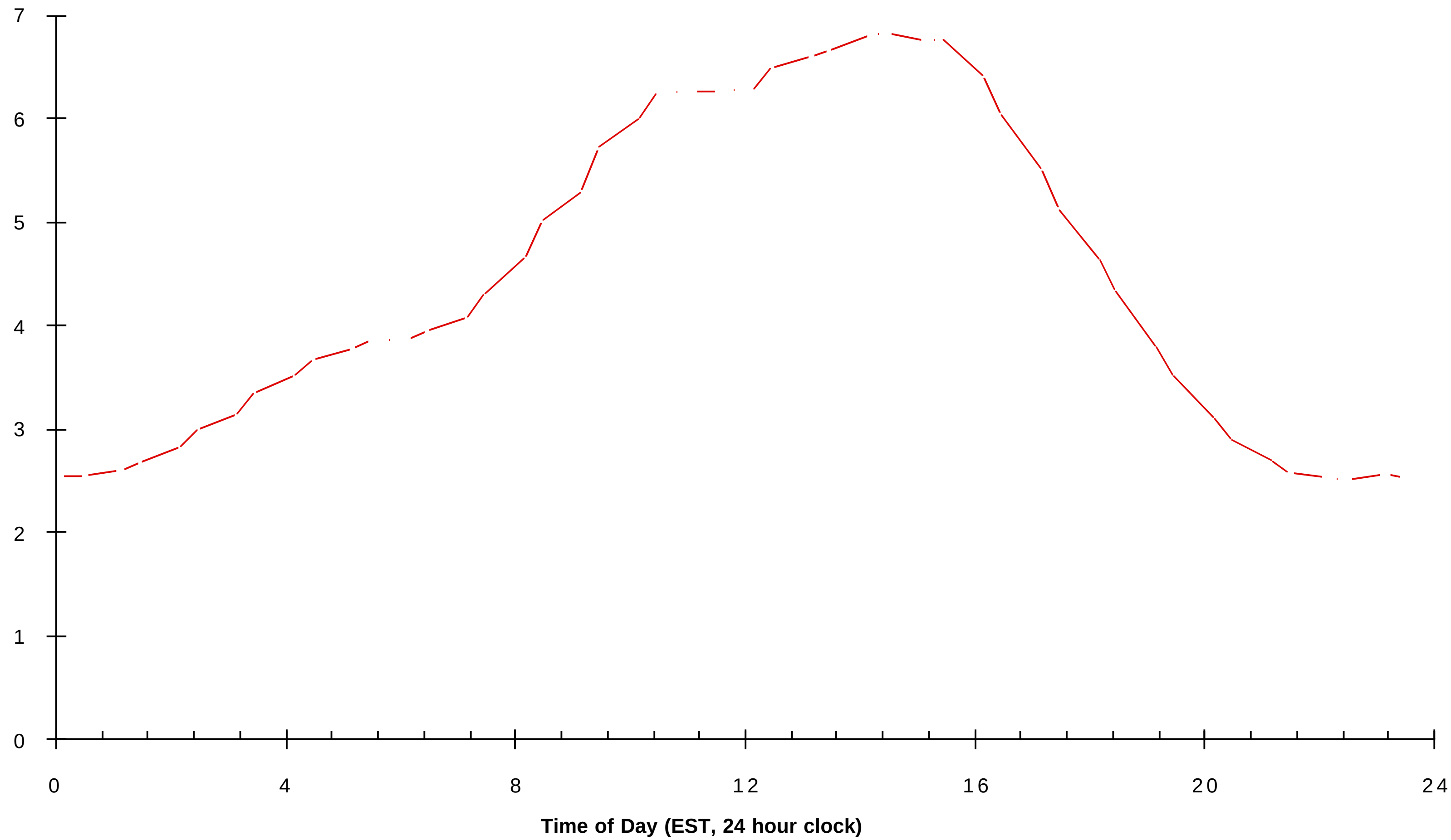
Function: Exponentially distributed interarrival times as a function of time of day.

Rationale: Studies of human behavior show that interarrival times are typically exponentially distributed. The number of requests per unit time is supplied by the User Model.

Notes: Graph on next slide gives the number of of requests (per unit time) as a function of time of day. This graph is for 3rd Quarter 1999.

This approach is at variance with requirement DADS3135, which specifies a lesser, constant load.

User Model gives a pull load that varies as function of time of day



Basis Functions (2)

Parameter: DAAC to which request is directed.

Function: Stochastic, probability per DAAC:

	ASF	EDC	GSFC	JPL	LaRC	NSIDC	ORNL
Probability	3.9%	48.4%	33.4%	1.7%	11.4%	0.6%	0.6%

Rationale: Best data available.

Notes: Probability developed from the User Model.

Basis Functions (3)

Parameter: Volume of data per request for ad hoc requests.

Function: Exponentially distributed request sizes, mean normalized to yield 2x volume distribution (when combined with user subscriptions and SCF/QA) from each DAAC.

Average request size = 260 MB.

Rationale: 2x volume distribution directed by ESDIS.

Notes: See Distribution Assumption (3). Normalization procedure is described on the next slide.

Normalizing procedure for the mean ad hoc request size

1. **Volume of data distributed per DAAC taken from User Model in baseline.**
2. **Separate volume into flow components.**
   - **subscriptions, DAAC integrity checks, and ad hoc retrieves**
   - **subscription volume based on assumption of percentage of total flow.**
3. **Calculate Archive-to-Distribution and Processing-to-Distribution traffic.**
4. **Run the dynamic model with only "push" enabled.**
   - **yields actual subscription volume based on products generated.**
5. **Run the dynamic model with all flows enabled.**
   - **user request size arbitrarily picked (typically 1GB by convention).**
   - **yields model archive-to-distribution and processing-to-distribtuion traffic.**
6. **Subtract subscription volume (from step 4) from both calculated traffic (from step 3) and model traffic (from step 5).**
7. **Normalized request size is the ratio of corrected calculated traffic (from step 6) to corrected model traffic (from step 6) multiplied by request size used in step 5.**

Basis Functions (4)

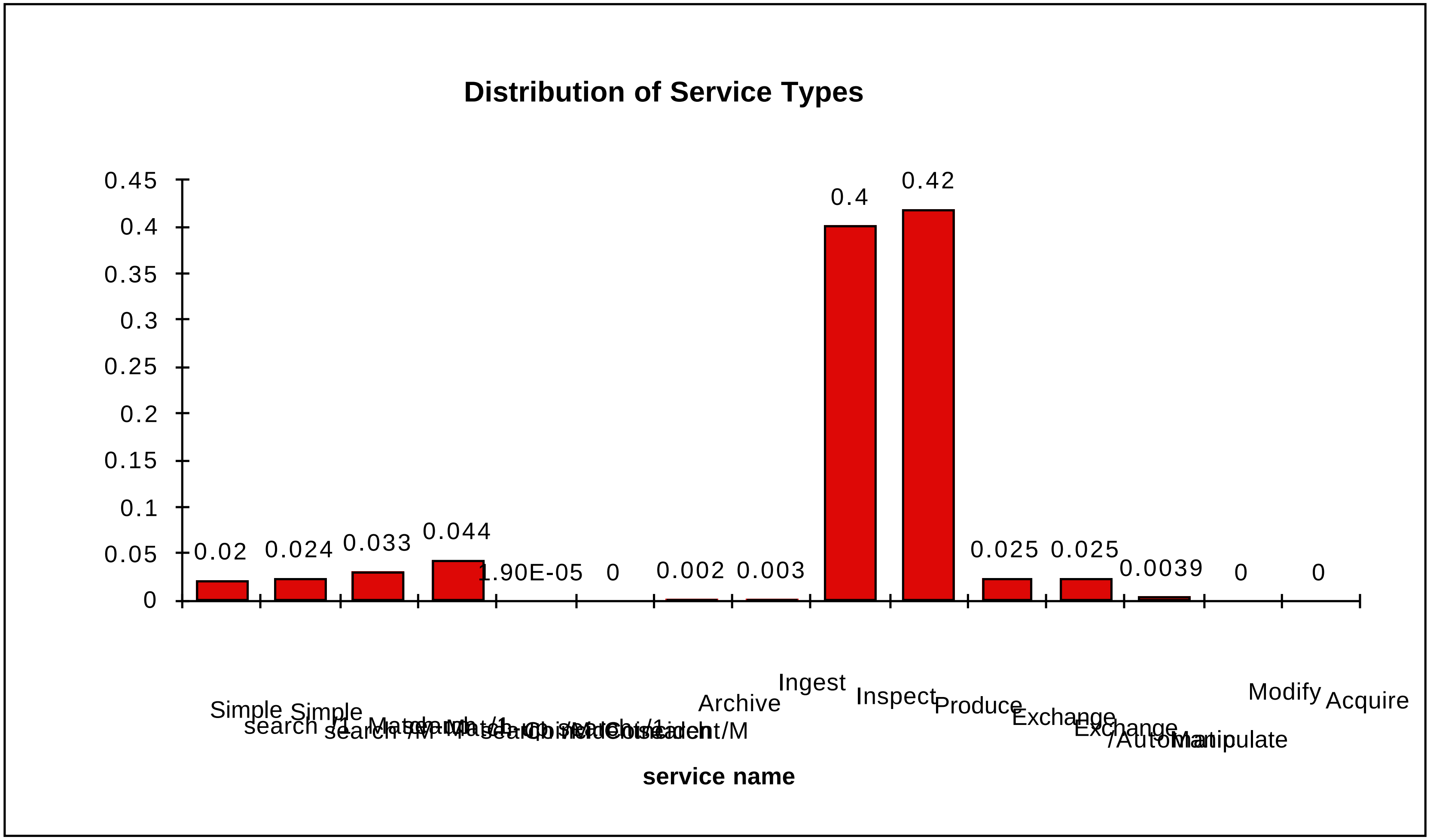
Parameter: Request type.

Function: Stochastic, probability per service given on next viewgraph.

Rationale: Best data available.

Notes: Probability developed from the User Model.

Distribution of Service Types



Basis Functions (5)

Parameter: Processing load due to manipulate requests

Function: Power law distribution.

Rationale: Engineering judgment

Notes: Power law distribution is similar to 80:20 rule; there is a large number of small requests and a few large requests.

Manipulate requests include requests requiring significant processing

- **certain subsetting requests**
- **on-demand processing**
- **user-supplied methods**

Basis Functions (6)

Parameter: Number of files input (from archive) during a manipulate

Function: Uniform random number between 1 and 10.

Rationale: Engineering judgment

Notes: None.

Basis Functions (7)

Parameter: Decimation ratio

Function: Parametric constant defaulted to 5-to-1

Rationale: Engineering judgment

Notes: The decimation ratio is ratio of the volume of data extracted from the archive to the volume of data distributed to the users during manipulate requests.

Basis Functions (8)

Parameter: Portion of data that is distributed electronically.

Function: stochastic, with probability of 50%.

Rationale: Interpretation of requirement DADS3110 when coupled with Assumption 10.

Notes: The other 50% is physical media.

Basis Functions (9)

Parameter: Portion of electronic distribution that is "pulled" (as opposed to "pushed").

Function: stochastic, with probability of 50%.

Rationale: Engineering judgment

Notes: None.

Basis Functions (10)

Parameter: CPU processing time for each PGE.

Function:

$$t_{cpu} = MFPOs_{process} / MFLOPS_{derated}$$

$$MFLOPS_{derated} = MFLOPS_{vendor} / factor_{derating}$$

Rationale: Definition of processing time.

Notes: CPU processing time is just one component of the time that a process "holds" a processor.

For $factor_{derating}$ see Processing Assumption (1).

Basis Functions (11)

Parameter: I/O time for each PGE.

Function:

$$t_{I/O} = (\sum^{all\ files} (\%\ read\ or\ written * size\ of\ file)) * factor_{latency}$$

$$factor_{latency} = (latency_{disk} / Block\ Size + 1 / Bandwidth_{I/O})$$

Rationale: Definition of time required to perform I/O.

Notes: I/O time is just one component of the time that a process “holds” a processor.

The latency calculation takes into average account sector and track latency.

Basis Functions (12)

Parameter: Elapsed processor time (also sometimes called "wall clock time").

Function:

$t_{wall} = \max(t_{I/O}, t_{CPU}) + (1 - \alpha) * \min(t_{I/O}, t_{CPU})$

where: α is the fractional overlap of I/O and Compute

Rationale: Definition of elapsed processor time.

Notes: See Basis Functions (11) & (12) for t_{CPU} and $t_{I/O}$.

Basis Functions (13)

Parameter: Archive read/write time.

Function: $t_{read/write}$ = size of file / $Bandwidth_{station}$

Rationale: Definition of archive read/write time.

Notes: $Bandwidth_{station}$ **is media dependent and is given in the Model Parameters section.**

Basis Functions (14)

Parameter: Exclusive (switched, non-shared) Network transfer time.

Function: $t_{exclusive}$ = size of file / $\text{Bandwidth}_{exclusive}$

Rationale: Definition of network transfer time.

Notes: Delays due to time spent in the queue are not included in this calculation. Since these delays are dependent on the instantaneous DAAC load, the model itself has to calculate the delay times.

Basis Functions (15)

Parameter: Shared Network transfer time.

Function: t_{shared} = (size of file / $Bandwidth_{shared}$) * Number in network

Rationale: Definition of network transfer time.

Notes: The model determines the number of files being transferred by the network at any given time.

Basis Functions (16a)

Algorithm to retrieve from or insert a file into the archive hardware

1. **Allocate both a robot and a read/write station**
2. **Cmd/Ack Delay to command the robot**
3. **Fetch to bring the tape from the archive to the read/write station**
4. **Mount to place the tape into the read/write station**
5. **Release the robot**
6. **Search to locate the desired file or position to write**
7. **Read or write the file**
8. **Rewind the tape**
9. **Allocate a robot**
10. **Cmd/Ack Delay to command the robot**

Basis Functions (16b)

11. Dismount the tape
12. Release the read/write station
13. Replace the tape back in the archive
14. Release the robot

Notes:

- **In step 1, the operation does not proceed until both are allocated.**
- **FSMS Delay is dependent on host computer and software and occurs before this set of operations. Therefore, it is not included in these steps.**

Basis Functions (17)

Parameter: Archive Tape Fetch/Replace Time

Function: Uniform random number over the range [0, $Fetch_{maximum}$]

Rationale: The desired tape could be immediately available to the robot or could be at the farthest range from the robot. This function simulates that while preserving the average robotic fetch delay.

Notes: $Fetch_{maximum}$ is dependent on silo manufacturer and is given in the Modeling Parameters section.

It is assumed that the replace time follows the same function as the fetch time.

Basis Functions (18)

Parameter: Archive Tape Search Time

Function: Uniform random number over the range [0, $Search_{maximum}$]

Rationale: The desired tape position could be anywhere within the tape. This function simulates that while preserving the average file position.

Notes: $Search_{maximum}$ is dependent on media type and is given in the Modeling Parameters section.

Basis Functions (19)

Parameter: Archive Tape Rewind Time

Function: Uniform random number over the range [0, $Rewind_{maximum}$]

Rationale: The current tape position could be anywhere within the tape. This function simulates that while preserving the average file position.

Notes: $Rewind_{maximum}$ is dependent on media type and is given in the Modeling Parameters section.

Basis Functions (20)

Parameter: Probability that a standard product granule (file) is subject to a subscription (standing order).

Function: stochastic, with probability of 50%.

Rationale: Engineering judgment

Notes: When combined with Distribution Assumption (3), this results in 25% of the volume of data distributed being due to subscriptions.

Basis Functions (21)

Parameter: Job scheduling algorithm.

Function:

- **On arrival of a file, each PGE which uses that file is checked and**
  - **(1)If any input file is non-extant, no action is taken.**
  - **(2)If all files are extant:**
    - **Files that are extant at the DPS are marked reserved so they will not be deleted.**
    - **Files that are non-extant at the DPS are staged from the DSS.**
  - **(3)If all input files are present at the DPS, the PGE is placed into the job queue**
    - **The job queue is a FIFO with priority per processing string.**

Rationale: Typical scheduling algorithm.

Notes: None.

Basis Functions (22)

Parameter: Time data is held on disk at DPS.

Function:

- **If a file is currently being used, it is left in place.**
- **If a file is marked reserved, it is left in place.**
- **If neither of these conditions exist, a timer is started. If when the timer expires the file is neither being used nor reserved it is deleted.**

Rationale: This disk management scheme is relatively easy to implement in the modeling tool. Keep in mind that each instantiation of a granule has to be tracked independently.

Notes:

- **Other disk management schemes will be investigated in the future.**
- **The current time-out period is 15 minutes.**

Basis Functions (23)

Parameter: Time data is held on disk at DSS.

Function:

If a file is not currently being used, a timer is started. If when the timer expires the file is neither being used nor reserved it is deleted. If the file is used during the time-out period, the timer is restarted.

Rationale: This disk management scheme is relatively easy to implement in the modeling tool. Keep in mind that each instantiation of a granule has to be tracked independently.

Notes:

- **Other disk management schemes will be investigated in the future.**
- **The current time-out period is set to 15 minutes.**

Basis Functions (24)

Parameter: Type of client access.

Function: In 3rd Qtr 1999 (epoch k), 90% of all client access will originate from WWW clients, 10% will originate from x-window (e.g., MOTIF).

Rationale: Engineering judgment

Notes: None.

Basis Functions (25)

Parameter: Percentage of requests referred to the V0 gateway.

Function: In 3rd Qtr 1999 (epoch k), 10% of client requests access V0 services. The remaining 90% access only ECS services.

Rationale: Engineering judgment

Notes: None.

Basis Functions (26)

Parameter: Single/Multiple DAAC query ratio.

Function: 55% of all user queries are resolved within a single DAAC, 45% require access of multiple DAACs.

Rationale: Data supplied by User Modeling.

Notes: Queries involving multiple DAACs will be handled by a DIM and the LIMs of the DAACs involved. Queries against a single DAAC are typically directed to the LIM of the DAAC involved.